

Sacherschließung als Basis für intelligente Navigation ausgehend von der DDC

Konzepte – Realisierung – Visionen

Lars G. Svensson

Über die Heterogenität der Sacherschließung in der deutschen Bibliothekslandschaft ist schon viel berichtet und geschrieben worden.¹ Im Folgenden wird dargestellt, wie eine Suchumgebung ausgehend von der DDC aufgebaut werden kann und wie diese auf andere Erschließungssysteme erweitert werden kann.

Klassifikatorische Suche mit der DDC

Einen ersten Einstieg in die Dokumentsuche mit der DDC bietet das hierarchische Browsing über die Klassifikation. Da davon ausgegangen werden kann, dass die Benutzer sich für Dokumente und nicht für Systemstellen interessieren, werden auf der Einstiegsseite die angeschlossenen Bibliothekssysteme (einzelne Bibliotheken oder Verbünde) zur Auswahl angeboten (Abb. 1).²

DDC-Suche

Bitte einen Bestand auswählen

[Deutsche Bibliothek Frankfurt am Main](#)

[Gemeinsamer Bibliotheksverbund](#)

[SUB Göttingen](#)

 svensson@dbf.ddb.de

[Lars G. Svensson](#) / 12.3.2004




Abb. 1

-
- 1 Vgl. z. B. Jürgen Krause [et al.]: Normierung und Standardisierung in sich verändernden Kontexten: Beispiel Virtuelle Fachbibliotheken. In: ZfBB 50 (2003) S. 19–28.
 - 2 Es besteht auch die Möglichkeit, direkt auf die Einstiegsseite einer Bibliothek oder eines Verbundes zu verlinken ohne über die allgemeine Einstiegsseite gehen zu müssen.

Wenn der Bestand ausgewählt worden ist – hier mit den Daten des GBV –, werden die Klassenbezeichnungen der DDC hierarchisch dargestellt (Abb. 2).

DDC-Suche: Gemeinsamer Bibliotheksverbund

Beschreibung	Treffer in dieser Kategorie	Treffer in dieser Kategorie und ihrer Unterkategorien
DDC	0 Titel	0 Titel
Computer, Information, allgemeine Nachschlagewerke	1 Titel	54519 Titel
Religion	3 Titel	30728 Titel
Naturwissenschaft	0 Titel	79148 Titel
Künste und Unterhaltung	0 Titel	58822 Titel

 svensson@dbf.ddb.de

Lars G. Svensson / 3.3.2004





   

Abb. 2

In der prototypischen Implementierung sind bisher die DDC-Bereiche 030-060 (Enzyklopädien, Zeitschriften, fortlaufende Sammelwerke, Verbände und Organisationen, Museumswissenschaft), 200 (Religion), 500-509 (Naturwissenschaften), 520-560 (Astronomie, Physik, Chemie, Geowissenschaften, Paläontologie) sowie 700 (Künste und Unterhaltung) eingespielt. Durch das Aktivieren der Hyperlinks können die Benutzer im Baum navigieren und somit von den allgemeinen Themenbereichen hin zu den spezielleren kommen. Parallel dazu wird angezeigt, a) wie viele Titel der gewählte Bestand besitzt, die genau der jeweiligen Systemstelle zugeordnet sind und b) wie viele Titel vorhanden sind in der Klasse und ihrer Unterklassen. In Abb. 3 z. B. wird angezeigt, dass der GBV 581 Dokumente hat, die genau der Systemstelle 720 (Architektur) zugeordnet sind (Suche nach ‚720‘) und dass er insgesamt 4444 Dokumente in der Klasse 720 und ihren Unterklassen hat (Suche nach ‚720*‘). Auf diese Weise sind die Benutzer immer über die zu erwartende Treffermenge informiert und können Suchanfragen, die entweder eine zu große (weil die Suchanfrage zu unspezifisch für das Thema wäre) oder gar keine (weil keine Dokumente vorhanden) Treffer liefern würden, vermeiden.

DDC-Suche: Gemeinsamer Bibliotheksverbund

Beschreibung	Treffer in dieser Kategorie	Treffer in dieser Kategorie und ihrer Unterkategorien
DDC	0 Titel	0 Titel
Künste und Unterhaltung	0 Titel	58822 Titel
Architektur	0 Titel	8577 Titel
Architektur	581 Titel	4444 Titel
Einzelne Aspekte der Architektur	0 Titel	0 Titel

svensson@dbf.ddb.de

[Lars G. Svensson](#) / 3.3.2004

Abb. 3

Neben den Klassenbezeichnungen bekommen die Benutzer auch Hinweise auf verwandte Begriffe, indem die Siehe- und Siehe-auch-Hinweise der jeweiligen Systemstelle angezeigt werden. Die referenzierten Systemstellen können per Hyperlink angewählt werden, was eine komfortable Navigation in der Klassifikation ermöglicht.

DDC-Suche: Gemeinsamer Bibliotheksverbund

Beschreibung	Treffer in dieser Kategorie	Treffer in dieser Kategorie und ihrer Unterkategorien
DDC	0 Titel	0 Titel
Künste und Unterhaltung	0 Titel	58822 Titel
Architektur	0 Titel	8577 Titel
Architektur	581 Titel	4444 Titel
<ul style="list-style-type: none"> • Für Stadtkunst und Landschaftskunst siehe Stadtkunst und Landschaftskunst 		
Philosophie und Theorie	149 Titel	265 Titel
Verschiedenes	1 Titel	389 Titel
Spezielle Themen der Architektur	11 Titel	227 Titel
Ältere Personen—Architektur für	0 Titel	0 Titel
Behinderte Personen—Architektur für	11 Titel	14 Titel
Historische, geografische, personenbezogene Behandlung	50 Titel	2791 Titel

 svensson@dbf.ddb.de

[Lars G. Svensson](#) / 3.3.2004



Abb. 4

Neben der hierarchischen Suche bietet das System auch eine verbale Suche über das DDC-Register an. Als Antwort auf die Suchanfrage (in unserem Fall „Kunst“) wird eine Resultatseite angezeigt (Abb. 5), die vom Aufbau her an das DDC-Register der gedruckten Ausgabe angelehnt ist.

DDC-Suche: Gemeinsamer Bibliotheksverbund

6 Treffer

- Kunst
 - [USA](#)
 - [angewandte](#)
 - [musikalische Repräsentation](#)
 - [religiöse Bedeutung](#)
 - [Christentum](#)

Abb. 5

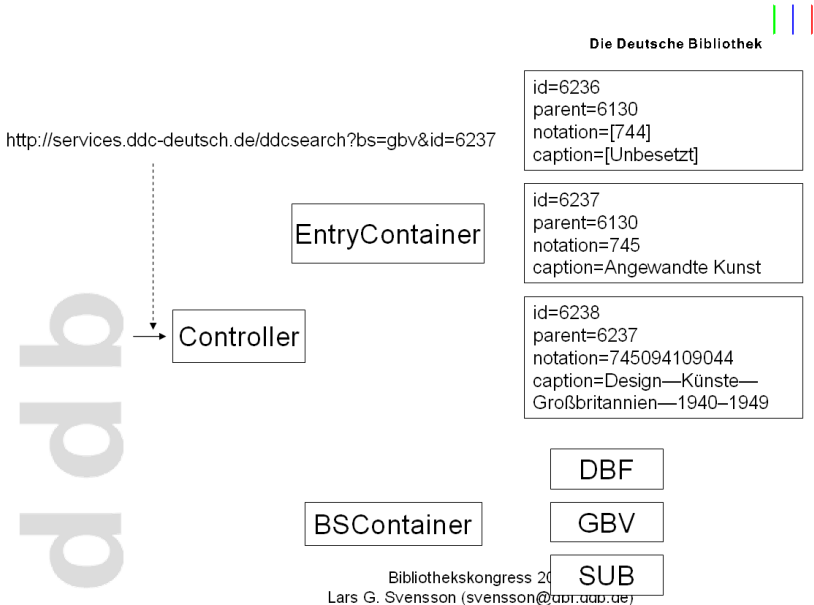
Da die Systemstellen in der Treffermenge über verschiedene Bereiche der Klassifikation gestreut werden können – in diesem Fall sind die ersten Treffer in der Kategorie 700 (Künste und Unterhaltung), während die letzten beiden auf Systemstellen im Bereich 200 (Religion) verweisen – ist es wichtig, dass die Benutzer früh entscheiden, unter welchem Aspekt sie das gewählte Thema behandeln wollen. Beim Aktivieren eines der Hyperlinks wechselt das System wieder in die hierarchische Ansicht, von wo aus weiter navigiert werden kann.

Wenn die gewünschte Systemstelle gefunden worden ist, entweder durch Browsing oder eine verbale Suche, können die Benutzer – sofern in der Systemstelle Dokumente vorhanden sind – einen der Links mit der Titelzahl aktivieren. Dabei verlassen sie das DDC-System und wechseln in den OPAC des gewählten Verbund- oder Lokalsystems.

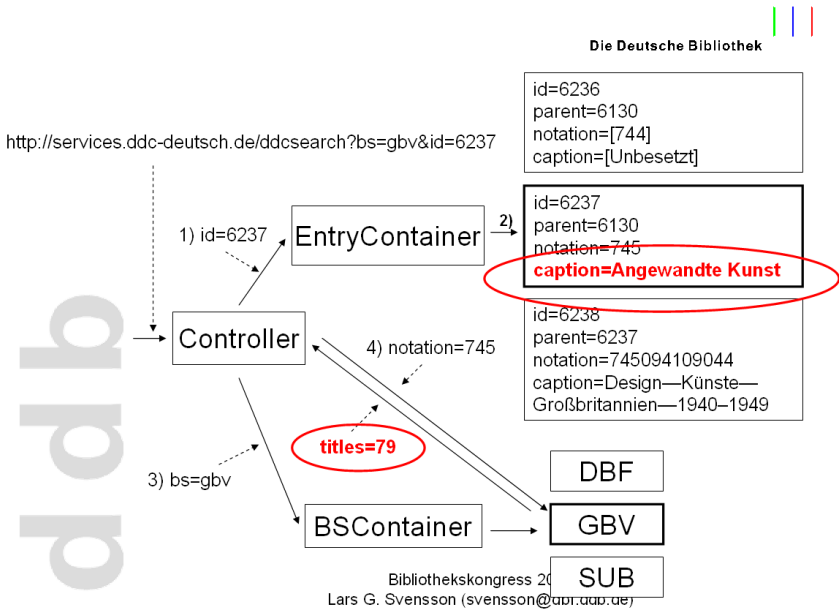
Technisches Konzept

Die eingehenden Anfragen an das DDC-System kommen zunächst zu einem RequestBroker, der den Ablauf steuert. Der Broker interagiert mit zwei weiteren Komponenten: einerseits mit einem EntryContainer, der als Schnittstelle zu der Datenbank mit den Klassifikationsdaten fungiert, andererseits mit den Schnittstellen zu den angeschlossenen Verbund- oder Lokalsystemen, hier gesammelt in dem BSContainer (BS=Bibliographic System). Die Klassen in dem EntryContainer besitzen u. a. eine eindeutige ID, eine Referenz auf die ID der übergeordneten Klasse, die DDC-Notation und die Klassenbezeichnung. Die bibliographischen Systeme dagegen sind abstrakte Schnittstellen, die auf das jeweilige Zielsystem zugeschnitten sind und standardisierte Zugriffsmethoden zum Abfragen der Titelzahl für eine bestimmte DDC-Notation und für

den Zusammenbau einer URL für den Wechsel in den OPAC des Zielsystems bieten. Die momentan existierende Implementierung modelliert den Zugriff auf ein PICA-System über die PSI-Schnittstelle, weitere Implementierungen für z. B. Aleph- und Horizon-Systeme sowie standardisierte Protokolle wie Z39.50 oder SRU sind geplant.



Als erstes identifiziert der RequestBroker anhand der Anfrage-URL das gewählte bibliographische System (bs=gbv) und erhält durch eine Anfrage an den BSContainer eine Referenz darauf. Danach liest er die ID der gewünschten Systemstelle aus (id=6237) und bekommt von dem EntryContainer die entsprechende Klasse zurückgeliefert. Als dritten Schritt nimmt der RequestBroker die Notation der Klasse und fragt das bibliographische System, wie viele Dokumente mit dieser Notation dort verzeichnet sind.



Dieses Verfahren wird für alle übergeordneten Systemstellen und für die unmittelbaren „Kinder“ wiederholt, danach wird aus den Trefferzahlen und den Klassenbezeichnungen die Resultatseite generiert.

DDC-Suche: Gemeinsamer Bibliotheksverbund

Suchen

Beschreibung	Treffer in dieser Kategorie	Treffer in dieser Kategorie und ihrer Unterkategorien
DDC	0 Titel	0 Titel
Künste und Unterhaltung	0 Titel	58822 Titel
Zeichnung, angewandte Kunst	0 Titel	5017 Titel
Angewandte Kunst	79 Titel	1524 Titel
Design, Künste — Großbritannien — 1940—1949	0 Titel	0 Titel
Antiquitäten	4 Titel	17 Titel
Gebrauchsgrafik und Industriedesign	127 Titel	239 Titel
Theoretisches und angewandtes Design, theoretische und angewandte Dekoration	110 Titel	423 Titel
Handwerkskunst	12 Titel	123 Titel
Kalligrafie, heraldisches Design, Illumination	4 Titel	385 Titel
Dekoratives Färben	1 Titel	30 Titel
Rundgemälde, Dioramen, Panoramagemälde	0 Titel	1 Titel

Suche nach Facetten

Das gerade vorgestellte System hat den Nachteil, dass es nur Zugriff auf diejenigen Notationsbestandteile bietet, die am Anfang der Notation stehen, und es nicht erlaubt, auf weitere Facetten aus den Hilfstafeln wie Geographica (Hilfstafel 2), ethnische Gruppen (Hilfstafel 5) oder einzelne Sprachen (Hilfstafel 6) zuzugreifen. So sind z. B. die Klassen 542.54 (Verfassungswesen in Indien) oder 639.5809483 (Schrimpfischen in Südwestnorwegen) in der hierarchischen Sicht nur über Verfassungswesen bzw. Schrimpfischen auffindbar. Eine Suche, die alle Dokumente über Skandinavien (T2—48) oder Indien (T2—54) findet, ist nicht möglich. Analoges gilt auch für z.B. sorbische Keramik (738.0899188) oder die Bibel in schottischem Gälisch (220.59163).

Eine Möglichkeit, nach Facetten zu suchen, wäre es, die Titeldatenbank mit Hilfe von regulären Ausdrücken zu durchsuchen. So würde der Perl-Ausdruck „\d{3}\.d*9163\d*“ (drei Ziffern, ein Punkt, null bis beliebig viele Ziffern, genau die Folge „9163“, null bis beliebig viele Ziffern) alle Notationen finden, die irgendwo nach dem Dewey-Punkt die Ziffernfolge ‚9163‘ haben und somit dem schottischen Gälisch zugeordnet werden könnten. Eine solche Suche wäre allerdings sehr ineffizient: Abgesehen davon, dass sie sehr langsam sein dürfte, da es sehr schwierig ist, einen Index dazu zu bauen und deshalb bei der

Suche jeder einzelne Titeldatensatz angefasst und überprüft werden müsste, würde sie auch einiges liefern, was überhaupt nicht zu dem Thema passt. So findet man mit „ $\{d\} \cdot d^{*54} d$ “ zwar alles, was Indien zugeordnet werden kann, aber auch z. B. 546.545 (Rhenium) oder 731.54 (Reliefplastiken). Man könnte zwar einwenden, dass für geographische Begriffe der Facettenindikator „09“, wie in 639.5809483, und für ethnische Gruppen der Facettenindikator „089“, wie in 738.0899188, vorgesehen ist und dass die korrekte Suchanfrage für Indien somit „ $\{d\} \cdot d^{*0954} d$ “ sein müsste, was grundsätzlich korrekt ist. Das Problem dabei ist jedoch, dass der Facettenindikator nicht immer vorhanden ist, wie das Beispiel indisches Verfassungswesen (542.54 statt 542.0954) zeigt.

Getrennte Ablage der Notationsbestandteile

Eine andere Möglichkeit, Facetten suchbar zu machen, wird im Rahmen des Projektes DDC Deutsch verfolgt. Der Ansatz liegt darin, nicht nur die gesamte DDC-Notation im Titeldatensatz abzulegen, sondern auch die einzelnen Notationsbestandteile. So würden für „Verfassungswesen in Indien“ neben der Gesamtnotation 542.54 auch die Einzelteile 542 (Verfassungswesen) und T2—54 (Indien) abgelegt werden. Für „Fischen nach Schalentieren in Südwestnorwegen“ wären die entsprechende Teile 639.5 (Fischen nach Schalentieren), 595.388 (Schrimps), T2—483 (Südwestnorwegen) und die Gesamtnotation 693.5809483.³ Eine Suchanfrage wie z. B. „schrimps skandinavien“ würde dann wie folgt ablaufen:

- 1) Suche im DDC-Register. Findet 595.388 (Schrimps) und T2-48 (Skandinavien)
- 2) Suche in der Titeldatenbank nach „ $ddc=595.388^* \text{ AND } ddc=T2-48^*$ “ (die Trunkierung ist wichtig, da auch Unterklassen mit einbezogen werden müssen). Unter den Treffern wäre auch der Datensatz oben, da wir genaue Übereinstimmung bei Schrimps (**595.388**) und – wegen der trunkierten Suchanfrage – Teilübereinstimmung bei Südwestnorwegen (**T2—483**) haben.

Mit dieser Vorgehensweise können die Notationsbestandteile einzeln indiziert werden, was zu einer Erhöhung der Suchgeschwindigkeit führt. Darüber hinaus werden auch Probleme mit dem so genannten Stehplatz gelöst.⁴ Da hier

3 Die Systemstelle 693.58 hat eine synthetische Notation, die sich aus 693.5 (Fischen nach Schalentieren) und der letzten ‚8‘ in 595.388 zusammensetzt.

4 Ein Stehplatz ist ein Thema, das kein ausreichend hohes Dokumentaufkommen vorweist, um eine eigene Notation zu erhalten, und das einen erheblich geringeren inhaltlichen Umfang hat als die Notation, bei der es aufgeführt wird. Standardschlüssel dürfen nicht an ein Thema in einer Stehplatz-Position angehängt werden; auch andere Formen der Notationssynthese sind nicht erlaubt.

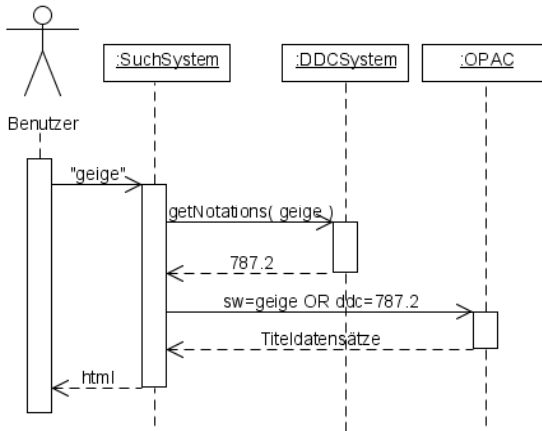
keine Anhängenzahlen verwendet werden dürfen, ist es unmöglich, anhand der Gesamtnotation nach Facetten zu suchen. So ist bei einer strikt nach DDC klassifizierten Arbeit über das Spielen mit elektrischen Eisenbahnen in Oklahoma der geographische Aspekt (Oklahoma) nicht sichtbar, da „Spielen mit elektrischen Eisenbahnen“ keine eigene Systemstelle hat, sondern Teil von 790.133 (Spielen mit Spielsachen) ist. Deswegen kann die Notation nicht durch die Anhängenzahl 09 (geographischen Schlüssel) + 766 (Oklahoma) zu 790.13309766 erweitert werden, was dazu führt, dass die Arbeit bei einer Suche nach Ressourcen, die sich mit Oklahoma beschäftigen, nicht gefunden werden dürfte. Indem die einzelnen Teilnotationen getrennt abgelegt werden, können sie aber trotzdem indexiert und suchbar gemacht werden und gleichzeitig kann die korrekte DDC-Notation – ohne Anhängenzahlen – für den bibliographischen Austausch bereitgestellt werden.

Thematische Suche

Um bei einer thematischen Suche sowohl die mit SWD als auch die mit DDC erschlossenen Ressourcen zu finden, muss ein Benutzer (mindestens) zwei Suchanfragen stellen: Eine an das DDC-System und eine an den gewählten OPAC. Dies führt dazu, dass er mit zwei verschiedenen Systemen zu tun hat, die zwei verschiedene Bedienungsoberflächen besitzen, und dass er zum Schluss die Literaturliste manuell zusammenstellen muss.

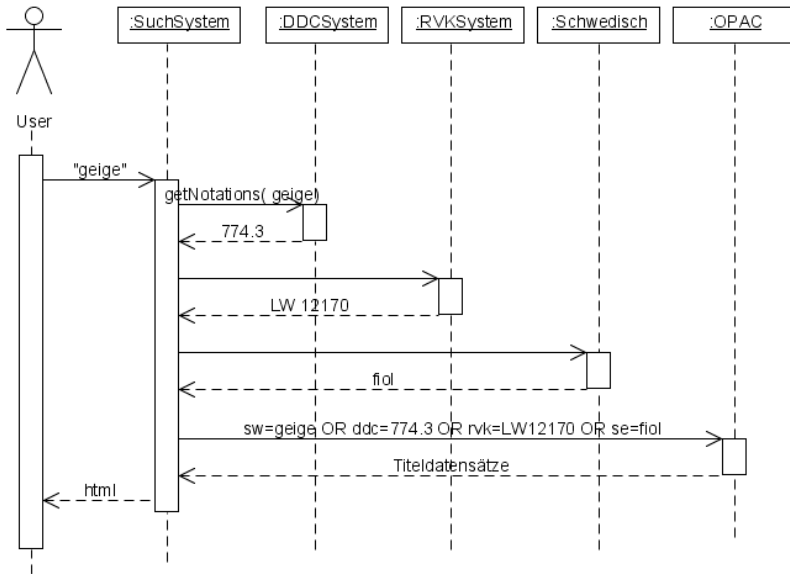
Es wäre deutlich eleganter, wenn man eine kombinierte Suche über DDC und SWD anbieten könnte, ohne dass der Benutzer viel davon mitbekommt. Im Folgenden soll der Versuch unternommen werden, die Grundstruktur eines solchen Systems zu beschreiben. Es sei darauf hingewiesen, dass die Beispiele natürlich stark vereinfacht sind.

Die Grundidee ist, die eingegebenen Suchbegriffe durch das Hinzufügen weiterer Suchtermini zu erweitern, wobei in unserem Beispiel ein DDC-Dienst dafür verantwortlich ist, für den gesuchten Begriff eine entsprechende DDC-Notation zu finden. Wenn der Benutzer z. B. nach „Geige“ (sw=geige) sucht, geht die Suchanfrage nicht sofort zum OPAC, sondern das Suchsystem sendet die Begriffe zuerst an das DDC-System und bekommt dafür eine oder mehrere (eventuell auch keine) DDC-Notationen zurück: in unserem Fall „787.2“, die Systemstelle für Geigen. Die ursprüngliche Anfrage wird dann durch eine ODER-Verknüpfung mit der erhaltenen Notation verbunden, so dass die Suchanfrage „sw=geige OR ddc=774.3“ an den OPAC geht.



Mit diesem Konzept lässt sich das System um weitere Dienste erweitern, die von einem Vokabular z. B. in RVK-Notationen, andere Thesauri oder andere Sprachen transferieren.⁵ So könnte man beispielsweise das System um eine RVK-Konkordanz und eine Crosskonkordanz Schwedisch erweitern. Jeder der Dienste übersetzt den Suchterm in sein Vokabular (idealerweise passiert dies parallel), und das Suchsystem setzt daraus eine Suchfrage zusammen, die an die Titeldatenbank gesendet werden kann.

5 Dies ist im Grunde das gleiche Konzept wie das der Transfermodule, wie es am IZ Sozialwissenschaften in Bonn entwickelt wird. Vgl. Jürgen Krause: Konkretes zur These, die Standardisierung von der Heterogenität her zu denken. In: ZfBB 51 (2004) S. 76–88.



Fazit: Navigation ausgehend von der DDC

Als das weltweit am meisten benutzte Klassifikationssystem hat die DDC ein erhebliches Potential, sich als Switching Language für sprachunabhängige Dokumentenrecherche zu positionieren. Allerdings ist eine verbale Suche für viele Benutzer deutlich intuitiver, so dass der Umstieg auf rein hierarchisches Browsing – wie es die DDC erlauben würde – nicht sinnvoll erscheint. Angesichts der großen Zahl an vorhandenen, mit SWD erschlossenen Daten wäre eine solche Vorgehensweise unsinnig. Vielmehr müssen die Bestrebungen darauf konzentriert werden, das Nebeneinander der verschiedenen Erschließungsmethoden auf der Ebene der Suchmaschinen zusammen zu führen. Für die wissenschaftliche Literaturrecherche birgt die Möglichkeit zur verteilten Suche über heterogen erschlossene Daten ein großes Potential. Die obigen Überlegungen stellen die Vorgehensweise dar, wie dies im Projekt DDC Deutsch verfolgt und realisiert wird.